Constructing Hierarchical Visual Tree for Discriminative Image Representation and Classification

Hao Lei[1], Ning Zhou[3], Kuizhi Mei[1], Peixiang Dong[1], and Jianping Fan[2]

[1] Institute of Artificial Intelligence and Robotics,
Xi'an Jiaotong University, Xi'an 710049, PR China
[2] School of Information Science and Technology, Northwest University,
Xi'an 710069, PR China
[3] Department of Computer Science, University of North Carolina,
Charlotte, NC 28223 USA

Abstract. To support large-scale visual recognition, it is critical to train a large number of classifiers with high discrimination power. To achieve this task, in this paper a hierarchical visual tree is constructed for organizing a large number of object classes and image concepts according to their inter-concept visual correlations. Based on the hierarchical visual tree, a novel approach is proposed for learning multi-scale group-based dictionary to support discriminative bag-of-visual-words (BoW-based) image representation. In addition, a structural learning approach is developed to enable large-scale classifier training over such hierarchical visual tree. We have also compared the performance of our hierarchical visual tree with traditional label tree over large-scale image collections.

Keywords: Hierarchical Visual Tree, Dictionary Learning, Discriminative Image Representation, Structural Learning, Image Classification.

1 Introduction

Image classification becomes increasingly important and necessary to support automatic image annotation, so that large-scale image retrieval can be made more intuitively by using the adequate keywords [23]. However, the performance of image classification largely depends on two inter-related issues: (a) discriminative representation for visual content of images, and (b) effective algorithm for multi-class classifier training.

For the first issue, bag-of-visual-words (BoW) has become one of popular methods for visual content representation of images due to its effectiveness and flexibility. BoW-based approach represents an image as a histogram based on the frequencies of a set of "visual words", which is known as visual dictionary [4]. Learning effective visual dictionary is a crucial issue for supporting discriminative BoW-based image content representation. For the second issue, when large-scale image categories come into view, large amount of classifiers should be effectively learnt for bridging the semantic gap successfully by mapping the

B. Huet et al. (Eds.): PCM 2013, LNCS 8294, pp. 637–648, 2013.

low-level visual features onto high-level image concept (human interpretation of image semantics).

In addition, when a large number of object classes (i.e., image semantics are interpreted by the visual content of object regions) and image concepts (i.e., image semantics are interpreted by the visual content of entire images) come into view, some of them are strongly inter-related (visually similar) because their relevant images may share some similar or even common visual properties (i.e., strong inter-concept visual correlations) [8,9]. In view of huge inter-concept visual correlations, to address the above two issues, robust techniques should leverage the inter-concept visual correlations for discriminative dictionary learning and effective large amount of classifiers training.

In this paper, a hierarchical visual tree has been constructed to organize a large number of object classes and image concepts according to their inter-concept visual correlations. We construct the hierarchical visual tree directly in the visual feature space rather than in the label space, because image representation and classifier training indeed happen in the visual feature space. On the one hand, we use the hierarchical visual tree to determine the groups of visual similar object classes and image concepts at different levels, and then jointly learn a discriminative dictionary for each group at different levels. On the other hand, we take advantage of the hierarchical visual tree to discover the inter-related classifiers for different object classes and image concepts, and develop a novel structural learning algorithm for effectively training the inter-related classifiers.

2 Related Work

In this section, we briefly review some related work on dictionary learning for BoW-based image representation and multi-class classifier training for image classification.

Most of prevailing algorithms have learnt a universal visual dictionary [17,21,27] for BoW-based image representation, where the same bases (the same set of visual words in the universal dictionary) are used to obtain the BoW histograms for all the images in the database. It is worth noting that the images from different object classes and image concepts may have diverse visual properties, thus such universal dictionary may not be optimum for all the object classes and image concepts. To improve the performance, Nistér et al. [18] developed a vocabulary tree to allow a larger and more discriminatory vocabulary, [14,28,30] learned the universal dictionary combined with the classifier training. However, these existing algorithms only considered intra-class (category) or intra-image visual correlations, while completely ignored inter-concept (category) visual correlations. Pioneering work proposed by Zhou et al. [31] have leveraged the inter-concept (category) visual correlations to learn multiple inter-related dictionaries jointly for the visually similar object classes and image concepts, where a common dictionary (which is shared among multiple visually similar object classes and image concepts) and multiple individual dictionaries (which are class-specific for each object class or image concept) are learnt jointly.

For multi-class classifier training, rather than providing an exhaustive and comprehensive review of all related works on image classification, we focus on giving a brief overview of the work on hierarchical image classification. In order to support hierarchical image classification, Barnard et al. [1], Vasconcelos et al. [24], and Fan et al. [8,9] have incorporated hierarchical mixture models and concept ontology to leverage the hierarchical inter-concept semantic similarity contexts for training multiple inter-related classifiers jointly. Fei-Fei et al. [12] have also incorporated prior knowledge of object parts and their locations to improve hierarchical image classification. The major problem with the hierarchical approach is that the classification errors may be propagated among the classifiers for the relevant image concepts (i.e., inter-concept error propagation) [10].

Some previous work have recently been proposed to characterize the inter-concept correlations for image classifier training. Marszalek et al. [15] and Fei-Fei et al. [5] have used WordNet [16] to find the semantic relationships between the labels and combined discriminative classifiers through the semantic hierarchies. Wang et al. [26] have utilized the normalized Google distance (NGD) [3] as the inter-word contextual potential for multi-label image annotation. However, these previous work measure the inter-concept correlations by using semantic information in the label space, it is not quite reasonable for determining the inter-related classifiers. Because classifier training and image classification are performed in the feature space rather than in the label space. Pioneering work proposed by Dong et al. [7] and Fan et al. [11] have utilized the inter-concept visual network for determining the inter-related classifiers and training them jointly.

3 Automatic Construction of Hierarchical Visual Tree

The images from the inter-related object classes and image concepts may share some similar or even common visual properties (e.g., visual features), i.e., these inter-related object classes and image concepts may have huge inter-concept visual correlations. In this paper, a hierarchical visual tree is constructed to organize the object classes and image concepts from multi-levels and determine the inter-related (visually similar) object classes and image concepts directly in the visual feature space.

To construct the hierarchical visual tree, we first propose a new method to characterize the inter-concept (category) visual correlations explicitly. We extract dense SIFT features at multi-scales for each image instance in the database, and then use spatial pyramid histograms [13] to represent each image instance. Let H_x and H_y respectively denote the spatial histogram descriptors of a pair of image instances x and y. A kernel function $k(x, y)$ is then defined to characterize their visual similarity relationship as follows

$$k(x, y) = \sum_{l=0}^{L} \alpha_l I_l(H_x^l, H_y^l) \tag{1}$$

Table 1. Inter-concept visual correlations

Concept pair	φ	Concept pair	φ	Concept pair	φ
crab & box turtle	0.96	monitor & microwave	0.62	canoe & shoji	0.38
deerhound & lynx	0.82	sky & cloud	0.62	bookcase & dune	0.37
go-kart & horse cart	0.78	clog & accordion	0.57	flat bench & sky	0.36
bookcase & cabinet	0.78	tiger cat & wave	0.51	microwave & crab	0.35
limousine & cab	0.74	cab & oilskin	0.44	lynx & website	0.33

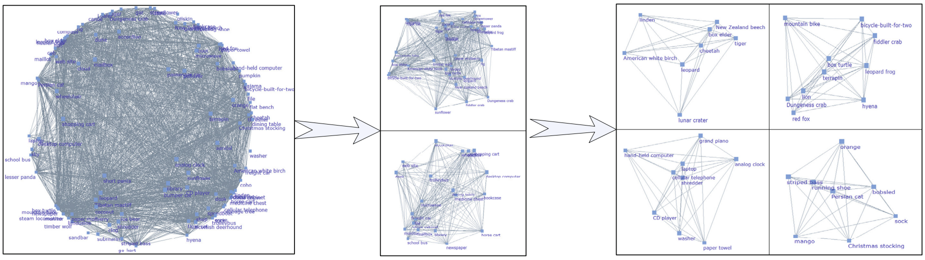
where $I_l(H_x^l, H_y^l)$ is the histogram intersection of H_x and H_y at pyramid level l, $l = 1, ..., L$ is the pyramid level, and α_l is the weight at level l, which is 2^{-L} for $l = 0$ and 2^{l-L-1} for others.

According to the image visual similarity relationship, for two given conceptual image sets C_i and C_j, we characterize their inter-concept visual correlation $\varphi(C_i, C_j)$ as follows:

$$\varphi(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{x \in C_i} \sum_{y \in C_j} k(x, y) \tag{2}$$

where $|C_i|$ and $|C_j|$ are the total numbers of image instances in the conceptual image sets C_i and C_j respectively, $k(x, y)$ is the visual similarity between two image instances x and y from conceptual image sets C_i and C_j. If relevant image instances from C_i and C_j share more similar visual properties (i.e., they are visually similar), their inter-concept visual correlation $\varphi(.,.)$ should have larger value. Table 1 gives some experimental results for the inter-concept visual correlations $\varphi(.,.)$ in our test dataset. It can be seen that our proposed method can well characterize the inter-concept visual correlations, e.g., relevant images from the image classes "bookcase" and "cabinet" look visually similar, and thus their inter-concept visual correlation has lager value in Table 1.

In the following, we use the top-down method to construct the hierarchical visual tree according to the inter-concept visual correlations. At the first level, we construct a visual concept network for all the object classes and image concepts in the database, where each object class or image concept is linked with multiple visually similar (inter-related) object classes and image concepts with larger values of the inter-concept visual correlations $\varphi(.,.)$. To determine the groups of visually similar object classes and image concepts at the first level, the visual concept network is treated as a weighted undirected graph, where each object class or image concept is treated as a node of the graph and the value of the inter-concept visual correlation $\varphi(.,.)$ is taken as the weight of the corresponding edge. And then, we partition this graph (visual concept network) into a set of disjoint subgraphs (visual concept sub-networks), and the Normalized cut (Ncut) [22] algorithm is employed to address this partition problem. These sub-networks constitute the second level of the hierarchical visual tree. Subsequently, we recursively do the above steps until reaching the maximum level L. Consequently, the hierarchical visual tree with L levels is constructed, which determines the multi-scale sub-networks of inter-related object classes and image



(a) Network at the first level

(b) Sub-networks at the second level

(c) Sub-networks at the third level

Fig. 1. Some visual concept networks at different levels of the hierarchical visual tree for our test dataset

concepts at different levels. Fig. 1 demonstrates some visual concept networks at different levels of the hierarchical visual tree for our test dataset (100 most popular object classes and image concepts from ImageNet [6] image set).

4 Multi-scale Group-Based Dictionary Learning

In Section 3, we have developed a new method for constructing the hierarchical visual tree directly in the visual feature space. The hierarchical visual tree is composed of a set of multi-scale visual concept sub-networks at the different levels, and each sub-network consists of a set of visually similar object classes and image concepts. We refer to such sub-network as a visually similar group. The feature vectors for such visually similar object classes and image concepts may overlap in the visual feature space, thus it is harder to distinguish them effectively. In this paper, we develop a group-based approach for leveraging the inter-concept visual correlations to jointly learn discriminative dictionary for all the object classes and image concepts in the same group. Our motivation is to learn a discriminative group-based dictionary for each group at different levels in order to enhance its discrimination power on distinguishing the visually similar object classes and image concepts in the same group, while different groups with weak inter-concept visual correlations will have different group-based dictionaries. In addition, because of different diversities of visual properties for different-level sub-networks, we learn different-scale dictionaries (i.e., multi-scale dictionaries) for different sub-networks (groups).

In the following, we discuss our algorithm for learning multi-scale group-based dictionary in details. We use the bottom-up approach to learn the dictionary for the visually similar groups at every level over the hierarchical visual tree. Because the inter-group visual correlations are weaker than the inner-group visual correlations at each level, so we can use the hypothesis that different groups at the same level are independent (i.e., ignoring weak inter-group correlations) in our algorithm. For legible expression, we first define some notations. Denote $l = 1, ..., L$ is the level of the hierarchical visual tree (first level is the top level and L-th level is the bottom level). Let G_i^l is the i-th ($i = 1, ..., M^l$)group at the l-th level, D_i^l is the dictionary corresponding to the group G_i^l, and the size of D_i^l is all K^l. First, we learn the dictionary for the groups at the L-th level (bottom level). Let X_i is the set of training samples from the object classes and image concepts in the group G_i^L, x_j, $j = 1, ..., N_i$ is a training sample from the set X_i, and w_k, $k = 1, ..., K^L$ be a visual word in the group dictionary D_i^L. We learn the $\{D_i^L\}$ by solving the following optimization problem:

$$\underset{\{D_i^L\}}{\arg\ \min} \sum_{i=1}^{M^L} \sum_{\substack{j=1 \\ x_j \in X_i}}^{N_i} ||x_j - \mathcal{V}(x_j, D_i^L)||^2 \tag{3}$$

where $\mathcal{V}(x_j, D_i^L)$ is the representation of the training sample x_j by using the group dictionary D_i^L. Here we utilize the vector quantization to accomplish the representation. Then $\mathcal{V}(x_j, D_i^L)$ is formulated as

$$\mathcal{V}(x_j, D_i^L) = \underset{w_k}{\arg} \min_{\substack{w_k \in D_i^L \\ k=1,\ldots,K^L}} ||x_j - w_k||^2 \tag{4}$$

We apply the iterative scheme to optimize the formula (3). At the beginning, we randomly select K^L training samples from each X_i to initialize the corresponding group dictionary D_i^L, and then we separately optimize each D_i^L iteratively. After we have learned the dictionary $\{D_i^L\}$ for the L-th level (bottom level), we then treat $\{D_i^L\}$ as the training samples and recursively learn the dictionary for other higher levels. It is worth noting that the parent groups from the high level only use the training samples from their children groups of the adjacent lower level. Because the more visual diversities for the group at higher level (such group consists of large amount of object classes and image concepts), we learn larger-scale dictionary for the group at higher level to enhance the discrimination power, while we choose the same scale for the the groups at the same level. As a result, the size of dictionary for the group at the l-th level is set to $2^{(L-l)} \times K^L$. Once multi-scale group-based dictionaries have been learned, we can use the BoW-based method to represent the images. It is worth noting that the children groups at low level are represented only by using the dictionary belong to their parent group at adjacent higher level, and images at different levels have different-scale representations.

5 Structural Classifier Training for Image Classification

In this section, we proposed a novel structural learning approach for effectively training multi-class classifiers over the hierarchical visual tree. We use support vector machine (SVM) with χ^2-based kernel [29] as the basic classifier to design our structural classifiers. Our structural classifiers consist of two components: intra-group classifiers for distinguishing the visually similar object classes and image concepts in the same group, and inter-group classifiers for discriminating those from different groups. For the intra-group, we only consider the groups at L-th level (bottom level), because the object classes and image concepts in such groups are most visually similar. We use the pairwise method to train the intra-group classifiers for these visually similar object classes and image concepts. For the inter-group, we treat the visually similar object classes and image concepts in the same group as a whole group concept at different levels. We learn inter-related classifiers for group concepts (image concepts for L-th level) in the same parent group at different levels, and only consider the adjacent high level groups from the same ancestry. Subsequently, we combine the intra-group and inter-group classifiers according to the hierarchical visual tree.

For a given object class or image concept C_j, X denotes the input test sample. We use $f(C_+, C_-, X)$ to denote the basic pairwise discriminant function, where C_+ denote the positive training concept(s) and C_- denote the negative training concept(s). The value of $f(C_+, C_-, X)$ is designed to 1 for the positive samples and 0 for the negative samples. Then our structural classifier $F(C_j, X)$ respect to $C_j (C_j \in G_i^L)$ can be defined as follows:

$$\begin{aligned} F(C_j, X) = & \overbrace{\beta_0 \sum_{C_k \in G_i^L, k \neq j} f(C_j, C_k, X)}^{intra-group} + \overbrace{\beta_1 \sum_{G_k^L \bowtie G_i^L, k \neq i} f(C_j, G_k^L, X)}^{bottom\ level\ inter-group} \\ & + \overbrace{\sum_{l=2}^{L} \beta_l \sum_{G_k^l \not\subseteq G_n^{l-1}} f(G_k^l, G_n^{l-1}, X)}^{high\ level\ inter-group} \end{aligned} \tag{5}$$

where $G_k^L \bowtie G_i^L$ denote the group G_k^L is inter-related with the group G_i^L; G_k^l and G_n^{l-1} respectively denote the group at l and $l-1$ level both including object class or image concept C_j. And β are the weighted coefficients, which determine the contribution ratio of the intra-group and inter-group classifiers to the final decision, and they are subjected to $\sum_{l=0}^{L} \beta_l = 1$. Once the structural classifiers have been trained, the automatic image classification is achieved by computing

$$C_{(X)} = arg \max_{C_j} F(C_j, X) \tag{6}$$

where $C_{(X)}$ is the predicted label for X. Fig. 2 demonstrates a simple example with 3 levels for our structural learning approach, where the intra-group, bottom level inter-group and high level inter-group are indicated expressly.

In the following, we discuss the computational complexity of our structural learning approach. Let N is the number of total object classes and image concepts. Without loss of generality, we suppose each group has M ($M \ll N$) children groups (M_i^L is the number of object classes and image concepts in the i-th group at level L). Therefore, the total number of binary classifiers for our structural approach is determined as follows

$$\sum_{i=1}^{M^L} \frac{M_i^L (M_i^L - 1)}{2} + N(M-1) + \sum_{l=2}^{L-1} M^l (M-1) \tag{7}$$

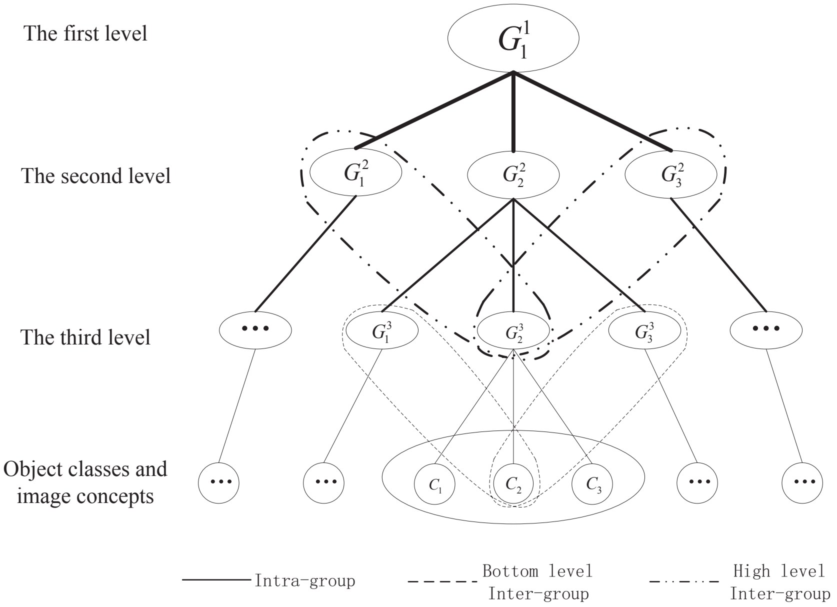


Fig. 2. A simple example for our structural learning approach

where the first item is the total number of intra-group classifiers, the second item is the total number of inter-group classifiers for the bottom level, and the third item is the total number of high level inter-group classifiers. Typically, $M^l \ll N$, $M \ll N$ and $M_i^L \ll N$, therefore our structural classifier training algorithm can achieve sub-quadratic complexity with the number of object classes and image concepts N, much less than the traditional pairwise method with the quadratic complexity of $O(N^2)$. Especially for large scale classification task (N is large), our structural learning algorithm is more efficient.

6 Algorithm Evaluation

6.1 Experimental Setup

Test Dataset: We chose the ImageNet [6] image set as our test dataset, because ImageNet is widely used for large scale image classification task and the images in ImageNet are more complex and have more inter-concept visual correlations. In our experiments, we selected 100 image concepts (most popular real-world object classes and scene categories) from different semantic levels and randomly selected 200 images for each image concept, which were used as the test bench for evaluating our algorithms. The selected subset from ImageNet image set is referred to as ImageNet100 in the following experiments. For ImageNet100 dataset, we randomly selected 100 images from each image concept as training data and the rest images as testing data.

Features Extraction: The images used in the evaluations are all preprocessed into gray scale, and we resize all the training and testing images to a max side length of 300 pixels (width or height) without changing their aspect ratios. SIFT descriptors are then computed at the points on a regular grid with spacing of 3 pixels for each image. In addition, in order to allow scale variation between images, multiple descriptors over supporting regions with different sizes (4, 6, 8, and 10 pixels) are computed at each grid point. We use the publicly available VLFeat toolbox (version 0.9.14) [25] to compute these SIFT descriptors.

Hierarchical Visual Tree: In the following experiments, we constructed the hierarchical visual tree with 3 levels according to our proposed inter-concept visual correlations for ImageNet100 dataset, and each parent group was partitioned into 4 children groups. It is worth noting that some groups at high level maybe stop growing in advance and become a leaf node of the visual tree.

Binary SVM Classifier: The SVM with χ^2-based kernel is used as the basic classifier, and the LIBSVM software [2] is employed to train the basic SVMs. For each basic SVM classifier, we use the same penalty parameter determined by searching over 7 values $(1, 2^2, ..., 2^{12})$ and the same kernel parameter determined by searching over 7 values $(2^{-9}, 2^{-7}, ..., 2^3)$ with five-fold cross validation on the training data.

6.2 Image Classification Evaluation

Performance evaluation for our algorithms and comparison with other algorithms: According to the algorithm proposed in Section 4, we have learned the multi-scale group-based dictionaries, and used the dictionaries to represent the images in ImageNet100 dataset. In the experiment, we evaluated three different sizes for K^L: 256, 512 and 1024. And then we trained the structural classifiers according to the hierarchical visual tree for multi-class image classification. We compared our algorithm with the traditional pairwise method and the hierarchical classification method over traditional label tree, the dictionary size used in these algorithms is set to 1024. We also compared our algorithm with some other inter-related classifier training algorithms. Fig. 3 demonstrates the classification accuracy of our algorithm and the hierarchical classification method [15] over all the object classes and image concepts on ImageNet100. The average classification accuracy for ImageNet100 dataset is shown in Table 2. From the experimental results, it can be seen that our algorithm outperforms all other compared methods. The reasons for this outcome are: (a) our algorithm can determine the inter-related object classes and image concepts in the visual feature space and effectively train the inter-related classifier for them; and (b) our structural learning approach can restraint the error propagation through the hierarchical tree.

Performance Comparison under various Measurements for Inter-concept Correlation Characterization: We have evaluated our correlation characterization approach compared with two popular measurement

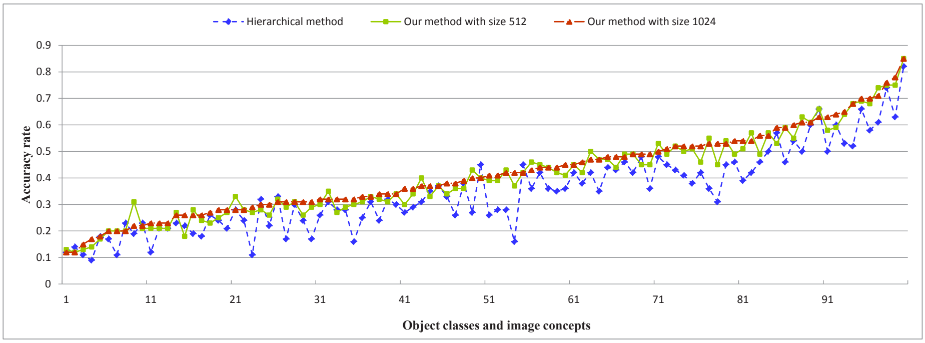


Fig. 3. Classification accuracy for our algorithm and the hierarchical classification method on ImageNet100 dataset

Table 2. Average classification accuracy comparison on ImageNet100 dataset

Algorithms	Average accuracy
Pairwise	0.4037
Hierarchical method [15]	0.3489
Inter-related method [7]	0.3696
Our method ($K^L = 256$)	0.3866
Our method ($K^L = 512$)	0.4031
Our method ($K^L = 1024$)	**0.4159**

Table 3. Inter-concept correlation measurement methods comparison on ImageNet100 dataset

Correlation Measurement	Dictionary Size (K^L)		
	256	512	1024
WordNet similarity [19]	0.3720	0.3833	0.3863
NGD [3]	0.3693	0.3836	0.3906
Our method	**0.3866**	**0.4031**	**0.4159**

methods which used WordNet distance and Google distance respectively. In our experiment, we used the method in [19] to compute WordNet distance between two object classes or image concepts, and the software tool provided by [20] (based on WordNet 3.0) was employed. Normalized Google distance (NGD) [3] is computed by incorporating the results of Google search engine to measure the inter-concept similarity between two object classes or image concepts. For given two image concept x and y, NGD for x and y is defined as follows

$$NGD(x,y) = \frac{\max\{\log f(x), \log f(y)\} - \log f(x,y)}{\log N - \min\{\log f(x), \log f(y)\}} \tag{8}$$

where $NGD(x,y)$ represents the normalized Google distance between the image concepts x and y. $f(x)$, $f(y)$ and $f(x,y)$ denotes the number of web pages containing x, y, both x and y, separately reported by Google. N is the total number of web pages indexed by Google. The experimental results of the average classification accuracy for different inter-concept correlation measurement approaches are shown in Table 3. One can observe that our inter-concept correlation measurement approach performs better than others.

7 Conclusions and Future Work

In this paper, a new method is developed for constructing the hierarchical visual tree automatically, where the inter-concept visual correlations are precisely characterized directly in the visual feature space. Based on the hierarchical visual tree, we propose a group-based approach for leveraging the inter-concept visual correlations to learn multi-scale dictionaries for discriminative image representation. In addition, a novel structural learning approach is also developed to effectively train inter-related classifiers for image classification. Our experiments over a large number of object classes and image concepts have approved the effectiveness of our algorithm for image classification task.

In the future, we will focus on how to determine the proper number of levels and groups at each level (i.e., the structure of the hierarchical visual tree). We believe that the more proper structure can provide better performance.

Acknowledgments. This work is partially supported by National Science Foundation of China under Grant No. 61272285, Ministry of education program for New Century Excellent Talents under NCET-11-0427, and Doctoral Program of Higher Education of China under Grant No. 20126101110022.

References

1. Barnard, K., Forsyth, D.: Learning the semantics of words and pictures. In: ICCV, vol. 2, pp. 408–415 (2001)
2. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011)
3. Cilibrasi, R., Vitanyi, P.: The google similarity distance. IEEE Trans. on Knowledge and Data Engineering 19(3), 370–383 (2007)
4. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: ECCV, vol. 1, p. 22 (2004)
5. Deng, J., Berg, A.C., Li, K., Fei-Fei, L.: What does classifying more than 10,000 image categories tell us? In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 71–84. Springer, Heidelberg (2010)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR, pp. 248–255 (June 2009)
7. Dong, P., Mei, K., Zheng, N., Lei, H., Fan, J.: Training inter-related classifiers for automatic image classification and annotation. Pattern Recognition 46(5), 1382–1395 (2013)
8. Fan, J., Gao, Y., Luo, H.: Integrating concept ontology and multitask learning to achieve more effective classifier training for multilevel image annotation. IEEE Trans. on Image Processing 17(3), 407–426 (2008)
9. Fan, J., Gao, Y., Luo, H., Jain, R.: Mining multilevel image semantics via hierarchical classification. IEEE Trans. on Multimedia 10(2), 167–187 (2008)

10. Fan, J., Luo, H., Hacid, M.S.: Mining images on semantics via statistical learning. In: ACM SIGKDD, pp. 22–31 (2005)
11. Fan, J., Shen, Y., Yang, C., Zhou, N.: Structured max-margin learning for inter-related classifier training and multilabel image annotation. IEEE Trans. on Image Processing 20(3), 837–854 (2011)
12. Fei-Fei, L., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: CVPR, vol. 2, pp. 524–531 (2005)
13. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: CVPR, vol. 2, pp. 2169–2178 (2006)
14. Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., Bach, F.R.: Supervised dictionary learning. In: NIPS, pp. 1033–1040 (2008)
15. Marszalek, M., Schmid, C.: Semantic hierarchies for visual object recognition. In: CVPR, pp. 1–7 (2007)
16. Miller, G.A.: Wordnet: a lexical database for english. Commun. ACM 38(11), 39–41 (1995)
17. Moosmann, F., Triggs, B., Jurie, F.: Fast Discriminative Visual Codebooks using Randomized Clustering Forests. In: NIPS, pp. 985–992 (2007)
18. Nistér, D., Stewénius, H.: Scalable recognition with a vocabulary tree. In: CVPR, vol. 2, pp. 2161–2168 (2006)
19. Patwardhan, S.: Incorporating dictionary and corpus information into a context vector measure of semantic relatedness. Master thesis, University of Minnesota, Duluth (2003)
20. Pedersen, T., Patwardhan, S., Michelizzi, J.: Wordnet: similarity - measuring the relatedness of concepts. In: AAAI, pp. 1024–1025 (2004)
21. Ries, C., Romberg, S., Lienhart, R.: Towards universal visual vocabularies. In: ICME, pp. 1067–1072 (July 2010)
22. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. on PAMI 22(8), 888–905 (2000)
23. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. on PAMI 22(12), 1349–1380 (2000)
24. Vasconcelos, N.: Image indexing with mixture hierarchies. In: CVPR, vol. 1, p. I (2001)
25. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008), `http://www.vlfeat.org`
26. Wang, Y., Gong, S.: Refining image annotation using contextual relations between words. In: CIVR, pp. 425–432 (2007)
27. Winn, J., Criminisi, A., Minka, T.: Object categorization by learned universal visual dictionary. In: ICCV, vol. 2, pp. 1800–1807 (October 2005)
28. Yang, L., Jin, R., Sukthankar, R., Jurie, F.: Unifying discriminative visual codebook generation with classifier training for object category recognition. In: CVPR, pp. 1–8 (2008)
29. Zhang, J., MarszaLek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. International Journal of Computer Vision 73, 213–238 (2007)
30. Zhang, Q., Li, B.: Discriminative k-svd for dictionary learning in face recognition. In: CVPR, pp. 2691–2698 (June 2010)
31. Zhou, N., Shen, Y., Peng, J., Fan, J.: Learning inter-related visual dictionary for object recognition. In: CVPR, pp. 3490–3497 (2012)